

DOCUMENT RESUME

ED 202 906

TM 810 350

AUTHOR Schaps, Eric; And Others
 TITLE A Process and Outcome Evaluation of an Affective Teacher Training Primary Prevention Program.
 INSTITUTION Pacific Inst. for Research and Evaluation, Napa, Calif.
 SPONS AGENCY National Inst. on Drug Abuse (DHEW/PHS), Rockville, Md.
 PUB DATE 25 Nov 80
 NOTE 59p.; For related document, see TM 810 349.
 EDRS PRICE MF01/PC03 Plus Postage.
 DESCRIPTORS *Classroom Environment; *Drug Abuse; *Elementary School Teachers; Formative Evaluation; *Inservice Education; Intermediate Grades; *Program Evaluation; Program Implementation
 IDENTIFIERS *Effective Classroom Management (Elementary)

ABSTRACT

Effective Classroom Management (ECM)-Elementary, an in-service course in which teachers were taught various communication, problem solving, and self-esteem enhancement techniques was evaluated. Process evaluation data included: (1) documentation of in-service training by outside observers; (2) teacher feedback on the individual training sessions; (3) questionnaire surveys of teachers at the end of training and the end of the school year; (4) interviews with a sample of teachers and principals; and (5) observations of classroom implementation. Process evaluation showed that the teachers highly regarded the training, used the skills competently in their classrooms, and found the communication and self-esteem enhancement (but not the problem solving) techniques useful. An analysis was conducted to determine whether the teachers' rates of implementing the techniques in the classroom were related to student outcomes; no such relationships were detected. (Author/GK)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED202906

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

A PROCESS AND OUTCOME EVALUATION OF AN AFFECTIVE
TEACHER TRAINING PRIMARY PREVENTION PROGRAM

SUBMITTED TO
NATIONAL INSTITUTE ON DRUG ABUSE
PREVENTION BRANCH

NOVEMBER 25, 1980

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

E. Schaps

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

The Napa Project
Pacific Institute for Research
and Evaluation
905 Jefferson Street
Napa, California 94558

Eric Schaps, Ph.D.
Joel M. Moskowitz, Ph.D.
John W. Condon, Ph.D.
Janet Malvin, Ph.D. Cand.

ABSTRACT

This study evaluated Effective Classroom Management (ECM)-Elementary, an in-service course in which teachers were taught various communication, problem solving, and self-esteem enhancement techniques. The goals of the course were to make classroom environments more responsible to students' affective and cognitive needs; thereby fostering positive attitudes, behaviors, and norms regarding self, peers, and school; and in so doing eventually reducing students' acceptance and use of psychoactive substances.

The experimental group consisted of the 23 4th-6th grade teachers in the four experimental schools and their 537 students. Eighteen of the 23 teachers were provided with ECM-Elementary in-service training. The 19 4th-6th grade teachers and 460 students in four control schools received no interventions.

The ECM in-service training consisted of twelve, weekly, two-hour, after-school sessions, supplemented by a minimum of four classroom visits by the trainer to observe and guide teacher implementation of the techniques covered in the training.

Process evaluation data included a) documentation of in-service training by outside observers, b) teacher feedback on the individual training sessions, c) questionnaire surveys of teachers at the end of training and the end of the school year, d) interviews with a sample of teachers and principals, and e) observations of classroom implementation. Process evaluation showed that the training was regarded very highly by the teachers, that the teachers used the skills competently in their classrooms, and that they found the communication and self-esteem enhancement (but not the problem solving) techniques quite useful. The teachers were found to vary considerably in how often they used most of the techniques in their classrooms, but none used the problem solving skills on a regular basis.

Experimental and control teachers and students were pre- and post-tested, and data regarding achievement and attendance were gathered from school district records. Teachers also rated their students' classroom behavior.

Post-test comparisons of experimental and control group teachers (adjusting for pretest differences) revealed no treatment effects on satisfaction with teaching, the importance and effectiveness of achieving teaching objectives related to the training, or on faculty cohesiveness.

The study employed a pre-post design with a control group. Eight elementary schools were paired based on characteristics of their student bodies, faculties, principals, and special programs. One school from each pair was assigned randomly to the experimental condition and the remaining schools served as controls.

Student- and class-level analyses of variance and covariance on outcome data revealed no treatment effects for boys. For girls, positive treatment effects were found on two variables. Experimental girls were less involved in marijuana use at the end of the school year than their control group counterparts, and they rated marijuana use as having greater adverse consequences. However, these results may be due to improperly specified models, attributable to the lack of pretest measures that better specify pre-existing student and teacher differences.

An analysis was conducted to determine whether the teachers' rates of implementing the techniques in the classroom were related to student outcomes. No such relationships were detected.

INTRODUCTION

Effective Classroom Management (ECM)-Elementary is an in-service course typical of much affective teacher training now being provided by primary prevention programs. In the ECM course, teachers are taught various communication, problem solving, and self-esteem enhancement techniques. They are also helped to apply the skills while teaching the regular academic curriculum. The goals of the course are to make classroom environments more responsive to students' affective and cognitive needs; thereby fostering positive attitudes, behaviors, and norms regarding self, peers, and school; and in so doing, eventually reducing students' acceptance and use of psychoactive substances.

The course curriculum is eclectic; it includes techniques developed by such innovators in education as Thomas Gordon (Active Listening; I-Messages), Louis Raths (Clarifying Responses), and Jack Canfield (various self-esteem enhancement activities). The curriculum does not directly address the issue of substance use; the only systematic discussion of substance abuse prevention in the training is a presentation on the links between substance abuse and various attitudes (e.g., low self-esteem, negative attitudes toward school) and behaviors (e.g., poor academic performance, high absenteeism). The course offers no information about psychoactive substances and gives no advice on how to handle substance abuse problems per se.

In recent years, many primary prevention programs have offered affective teacher training courses similar to ECM. Several state-wide networks of programs offer such training, including Michigan (Michigan Department of Education, Note 1), Georgia (Georgia Department of Human Resources, Note 2), and Pennsylvania (Bandt et al, Note 3). Affective teacher training programs have been shown effective at improving peer relationships, classroom discipline, attitudes toward school, and academic achievement (Baskin and Hess, 1980). A recent review of prevention program evaluations suggested that training "significant others" (e.g., parents teachers) in affective skills may be an effective approach to prevention (Schaps et al, in press).

In the present study, conducted during the 1978-79 school year, ECM-Elementary training was provided to 18 4th, 5th, and 6th grade teachers in four elementary schools. The teachers applied the in-service skills in their classrooms under the continuing guidance of the ECM trainers. The teachers, then, were the critical links in the delivery of this intervention; an adequate "treatment" depended upon the teachers using the in-service skills appropriately and consistently. Consequently, the present study included an intensive process evaluation, designed to monitor the teachers' reactions to the training and their use of the ECM skills.

Outcome evaluation in the present study assessed the impact of ECM on both the teachers and their students. Three teacher outcome variables were measured: the perceived importance of, and effectiveness at, achieving the objectives of the ECM training; satisfaction with teaching; and perceived cohesiveness of each school's faculty. We hypothesized that ECM would impact all three of these outcome variables.

Student outcome measures included locus of control, social self-esteem, academic self-esteem, attitudes toward peers, perceived classroom climate, attitudes toward school, perceived peer attitudes toward school, total absences, unexcused absences, reading achievement, math achievement, and misbehavior in school. Drug-specific outcome measures included perceived utilities of tobacco, alcohol, and marijuana use; and involvement in use of each of these substances. Among these variables, we hypothesized that ECM would be most likely to impact perceived classroom climate, attitudes toward school, and social self-esteem because the in-service skills bear most directly on these variables. We also anticipated that ECM would impact other variables, although it does not address the others as directly. Among these other variables, we expected that the drug-specific measures would be least likely to show an effect, due to the low level of drug involvement at these ages and the relatively brief duration of the intervention. (The present study will be continued for two additional years, with parts of each of the comparison groups receiving additional interventions each year.)

METHOD

Assignment of Schools to Condition

Eight elementary schools (grades K-6) from a predominantly white, middle-class, suburban public school system in Northern California were paired based on characteristics of their students, faculties, principals and special programs. One school from each pair was randomly assigned to the experimental condition and the other to the control condition.

Table 1 shows, for the experimental and control conditions, the school means and standard deviations of the measures employed in matching the schools. For each school, these measures included: a) total student enrollment in the spring prior to the study; b) the number of students in grades 3-5 during the prior spring; c) the number of "pure" grade 4-6 classes and the number of "mixed" classes containing students of different grade levels; d) average reading scores for students in grades 3-5 on the Stanford Achievement Test administered during the prior spring; e) the percentage of students from families receiving Aid for Families with Dependent Children (AFDC); f) the percentage of students classified as Spanish-speaking with either limited or no English language skills; g) the average number of unexcused absences per student during the prior school year; and h) an estimate of the student body's socioeconomic status (made by a school district official).

TABLE 1

COMPARISON OF EXPERIMENTAL AND CONTROL SCHOOLS AFTER
MATCHING AND RANDOM ASSIGNMENT TO CONDITION

	<u>Experimental Schools (N=4)</u>		<u>Control Schools (N=4)</u>	
	<u>School Mean</u>	<u>Standard Deviation</u>	<u>School Mean</u>	<u>Standard Deviation</u>
Total School Enrollment	347	179	282	98
Grades 3-5 Enrollment	152	85	127	48
Number of 4-6 Grade Classes per School				
Pure grade classes	3.5	3.3	4.5	1.3
Mixed grade classes	2.3	1.3	.3	.5
SAT Reading Achievement Scores	81.8	3.9	80.5	3.7
Percent of Enrollment Receiving AFDC	8.3%	9.9%	5.5%	2.4%
Percent of Spanish-Speaking	0.5%	0.8%	1.4%	2.7%
Average Number of Unexcused Absences per Pupil	1.2	0.1	1.8	.6
Socioeconomic Status Rating	54	17	59	18
Eligible Faculty Support Rating	.72	.33	0.54	.16
Eligible Faculty Competence Rating	.30	.42	.44	.19
School Principal Support Rating	.75	.29	.75	.29
School Principal Competence Rating	.13	.25	.63	.25

In addition, two district officials rated each school regarding the degree to which: a) eligible teachers would support and participate in the in-service training; b) eligible teachers were already competent in classroom management and interpersonal skills; c) the principal would support the in-service training; and d) the principal had influence over his or her teaching staff.

The experimental and control schools were similar on 11 of the 13 measures used in the assignment to treatment condition. Experimental schools did contain more mixed grade classrooms ($t = 2.87$, $p < .05$) and their principals generally received lower competence ratings ($t = -2.83$, $p < .05$). Those two initial differences were considered not to constitute a substantial bias to the design.

Subjects

All grade 4-6 students and their 23 teachers in the four experimental schools constituted the experimental group. Eighteen of the experimental teachers completed the ECM in-service training.¹ The control group consisted of all grade 4-6 students and their 19 teachers in the four control schools.² There were 21 classrooms in the experimental group: 13 single grade classes and 8 mixed classes. The control group contained 18 classrooms: 15 single grade classes and 3 mixed classes.

At the beginning of the study, 574 students were enrolled in grades 4-6 of the experimental schools and 497 were enrolled in the control schools. Thirty-seven students were excluded from each condition due to lack of parental permission for testing or because they transferred out of the school during

¹Four of the participating teachers taught half-time; thus, 16 classes had participating teachers. Of the nonparticipating experimental teachers, one taught grade 4; one, grade 5; two, grade 6; and one, grades 5-6.

²Two of the control teachers taught half-time.

the year. The experimental group consisted of 291 males and 246 females³, and the control group consisted of 215 males and 245 females. The ethnic composition of the sample was 94% ($N = 937$) White and 6% ($N = 60$) minority, with Mexican-Americans ($N = 34$) and Asian/Pacific Islanders ($N = 18$) comprising the largest minority groups.

In-Service Training Program

The ECM in-service program consisted of 12 weekly, two-hour, after-school training sessions. These were held between October, 1978 and January, 1979. The first and second sessions were mandatory for the 23 grade 4-6 teachers in the four experimental schools. These first two sessions were conducted on "minimum days," which were granted by the school district in order to ensure the teachers' exposure to the initial training, and thereby interest them in the remaining voluntary sessions.

The teachers were offered a \$200 stipend for attending the voluntary sessions regularly and making a sincere effort to apply the in-service skills in their classrooms. They were also offered the opportunity to obtain graduate-level credit from a local university for completing the training.

One of the trainers observed the teachers applying the skills in their classrooms on a minimum of four occasions. These visits began during the training and continued until May, 1979. After each visit, the trainer met briefly with the teacher to provide feedback.

³71 males and 55 females were enrolled in classrooms with nonparticipating experimental teachers.

The trainers had five and eight years respectively of experience conducting in-service programs for teachers. Both also had previous experience as classroom teachers and as school administrators. One had taught many of the ECM skills in previous training efforts.

Many of the training sessions followed a standard format. First, a self-concept enhancement technique or activity was introduced. Then the skills learned in prior sessions were reviewed at length, along with the teachers' experiences using the skills during the previous week. Finally, a new skill was introduced and practiced in various ways.

Five communication skills were taught during the first six training sessions:

- Tailored Feedback--a technique for accurately and nonjudgmentally describing another's behavior and its consequences;
- I-Messages--a technique for effectively communicating about the impact of another's behavior upon oneself;
- Clarifying Responses--brief questions and comments by which a listener can prompt further thinking by a speaker;
- Reflecting Feelings--techniques which enable a listener to indicate acceptance and accurate understanding of the speaker's feelings; and
- Reflecting Content--techniques which enable a listener to show interest and to indicate understanding of the speaker's message.

Five problem solving skills were taught during the last six sessions of the training:

- Creating Problem Statements--a procedure and format for clarifying a problem and breaking it into component parts;
- Brainstorming--a procedure for generating potential responses to a problem or opportunity;
- Evaluating Alternative Solutions--procedures for assessing the advantages and disadvantages of potential solutions;

- Force Field Analysis--a procedure for identifying factors which will promote and hinder particular courses of action; and
- Developing an Action Plan--a procedure and format for specifying a plan of action to solve a problem.

Self-concept enhancing activities, exercises and techniques were introduced at six of the sessions. In most of these, each participant was given the opportunity to describe and receive group recognition for a positive skill, trait, or experience.

Student self-report outcome data. Student pretest and post-test self-report data were obtained with the Self Observation Scales (Intermediate Level, Form C) developed by Stenner and Katzenmeyer (Note #4). This nationally normed instrument has empirically determined scales which measure the ways children perceive themselves and their relationships to their peers, their teacher, and their school. The questionnaire contains 60 statements to which students respond "yes" or "no."

Pretest data were also obtained with the Intellectual Achievement Responsibility Questionnaire (Crandall, Katkovsky, and Crandall, 1965). Four items referring to "parents" were adapted to read "an adult who knows you" in order to conform to California Education Code. This questionnaire measures the belief in one's own control over, and responsibility for, intellectual-academic successes and failures.

On the post-test, data were obtained from students with the Student Questionnaire (ELE-2), an instrument developed for this study. This instrument consists of two sections. Part 1 contains 18 items selected from the Intellectual Achievement Responsibility Questionnaire based upon

published psychometric properties. Part 2 contains three sets of items:

a) the Scholastic subscale from the Intermediate-Level of the Self Appraisal Inventory (20 items) developed by the Instructional Objectives Exchange (1972b), a criterion-referenced measure of academic self-esteem; b) the Authority and Control (12 items) and Interpersonal Relationships With Pupils (12 items) subscales from the Intermediate-Level of the School Sentiment Index developed by the Instructional Objectives Exchange (1972a), a criterion-referenced measure of attitudes toward school; and d) a measure of perceived peer attitudes toward school developed for this study by adapting 11 items from eight instruments that measure attitudes toward school.

Drug-related post-test data were collected using the My Opinion Survey (E1-2). This instrument was developed for this study to assess for each of the three "gateway" drugs, alcohol, cigarettes, and marijuana: the students perceived costs and benefits for use of the drug, their lifetime use, their current use ("last four weeks"), their attitudes toward use, their perceptions of peers' attitudes toward use ("most kids in my class"), and their perceptions of peer use. In addition, a ten-item inferential scale that measures predisposition to use patent medicines was adapted from the Life Decisions instrument (Instructional Objectives Exchange, 1973).

The final measures employed in the data analysis for the present study were derived from a theoretical model and empirical scaling analyses. The details of the scaling procedures and results have been reported by Moskowitz, Condon, Brewer, Schaps, and Malvin (Note #5). Pretest and post-test data were collected from a large sample of students in grades 3 and 4, in addition

to the students in this study. Random samples were selected from each grade level. Item means, variances, and intercorrelations were found to be similar across grade levels; thus, scaling procedures were applied across grades. Item intercorrelations from the pretest and from the post-test data were subjected to separate multiple group confirmatory factor analyses.⁴ The resultant pretest and post-test scales appear in Table 2 with the number of items contained in each, and their internal consistency reliabilities estimated by coefficient alpha. The final scales included pretest and post-test measures of locus of control for success, locus of control for failure, social self-esteem, attitudes toward peers, affective teaching climate, and attitudes toward school. The measures of social self-esteem, attitudes toward peers, and attitudes toward school employed the same items in the pretest and post-test. For the locus of control measures, a subset of pretest items was used for the post-test, and for affective teaching climate, a subset of post-test items was used for the pretest. Other post-test measures included academic self-esteem, perceived peer attitudes toward school, perceived costs of and involvement in alcohol, cigarette and marijuana use. The involvement scales consisted of use, attitude, and perceived use and attitude items.⁵ The measures of perceived benefits of drug use and predisposition to use patent medicines were eliminated due to inadequate psychometric properties.

⁴The matrices were computed using pair-wise deletion of missing values and communalities were inserted into their diagonal elements.

⁵The involvement scales measure perceptions of peer attitudes and peer use as well as own attitudes and use. For each substance, the five component items exhibited good internal and external consistency, providing strong empirical support for constructing a single unidimensional construct. The positive relationships obtained between the perceived peer measures with own use and attitude measures indicate that students' own use and attitudes parallel their peers' or that they project their use and attitudes onto their peers.

TABLE 2

14.

STUDENT PRETEST AND POST-TEST SCALES, NUMBER OF ITEMS AND
INTERNAL CONSISTENCY RELIABILITIES (COEFFICIENT ALPHA)

Subscale	PRETEST			POST-TEST		
	N	Number of Items	Reliability	N	Number of Items	Reliability
Locus of Control: Success (Control Suc)	1040	12	.64	513	7	.56
Locus Of Control: Failure (Control Fail)	1040	10	.65	513	7	.62
Social Self-Esteem (Social Self)	517	6	.62	513	6	.66
Attitudes Toward Peers (Att Peers)	517	8	.76	513	8	.80
Affective Teaching Climate (Affec Climate)	517	7	.74	513	17	.91
Attitudes Toward School (Att School)	517	6	.68	513	6	.74
Academic Self-Esteem (Acad Self)	--	NA ^a	--	513	11	.79
Perceived Peer Attitudes Toward School (Peer Att Sch)	--	NA	--	513	8	.72
Perceived Costs of Alcohol Use (Alc Costs)	--	NA	--	386	5	.70
Perceived Costs of Cigarette Use (Cig Costs)	--	NA	--	386	5	.63
Perceived Costs of Marijuana Use (Pot Costs)	--	NA	--	386	5	.67
Involvement in Alcohol Use (Alc Involve)	--	NA	--	386	5	.77
Involvement in Cigarette Use (Cig Involve)	--	NA	--	386	5	.75
Involvement in Marijuana Use (Pot Involve)	--	NA	--	386	5	.84

^aNA indicates that this scale was not administered

Student archival outcome data. In addition to the self-report data gathered directly from students, we obtained achievement data collected by the school district as part of its regular testing activities. We employed the total reading (Read) and total mathematics (Math) stanine scores from the Stanford Achievement Test, Intermediate Levels I (grade 4) and II (grades 5 and 6) (Madden, Gardner, Rudman, Karlson, and Merwin, 1973), administered in May, 1979.

Two different measures of student attendance were taken. The total number of unexcused absences for the school year (Unex Abs) was obtained for each student from the school district's records. This type of absence occurs when a student is absent from school and does not provide the school with a parental excuse indicating that the student was sick. In addition, the average monthly number of absences for each student (Total Abs) during the fifth through eighth school months (January through April) was extracted from teachers' records by project staff.

Students' sex and ethnicity were determined from school district records.

Teacher data on student behavior. Teacher ratings of student misbehavior were obtained at the time of the post-test with the Student Behavior Report. This instrument provided each teacher with the class roster and asked the teacher to indicate how frequently each child had been minor (Minor) and major (Major) discipline problems during the past four months (January through April) on a five-point scale ranging from "never" to "about once a day or more."

Teacher outcome data. Teacher pretest and post-test self-report data were obtained with different versions of the Teacher Questionnaire, an instrument developed for this study. This instrument included measures of a) satisfaction with teaching (Teacher Satisfaction) adapted from the Purdue Teacher Morale Inventory (Rempel and Bentley, 1964); b) faculty cohesiveness adapted from the Teacher Cooperation Subscale of the Teacher Attitude and Classroom Climate Questionnaire (Kaufman, Semmel and Agard, Note 6) and from the Intimacy subscale of the Organization Climate Description Questionnaire (Halpin and Croft, 1963); and c) the importance and effectiveness of achieving teaching objectives related to the in-service training (ECM objectives), a measure developed for this study. The post-test also included items developed for the study comprising an inventory of classroom behaviors relevant to the ECM in-service training. The number of items included in all pretest and post-test scales and their internal consistency reliabilities estimated by coefficient alpha are shown in Table 3 for grades 3-6 teachers.⁶ The classroom behavior inventory items were not employed in the outcome analysis since the data generally failed to exhibit the expected correlations with the classroom observation data and the teacher skill implementation data in the sample of teachers who participated in the training. However, a subset of these items was used in a treatment implementation analysis described later.

⁶To increase stability, data from additional grade 3-4 teachers were included in the scaling analysis.

TABLE 3

TEACHER PRETEST AND POST-TEST SCALES, NUMBER OF ITEMS, AND
INTERNAL CONSISTENCY RELIABILITIES (COEFFICIENT ALPHA)

<u>Scale</u>	<u>Pretest</u>			<u>Post-test</u>		
	<u>N</u>	<u>Number of Items</u>	<u>Reliability</u>	<u>N</u>	<u>Number of Items</u>	<u>Reliability</u>
Teacher Satisfaction	73	3	.68	60	8	.80
Faculty Cohesiveness	73	7	.87	60	7	.90
ECM Objectives	73	6	.74	60	6	.76

Data Collection Procedures

Student survey. The student pretest was administered in October, 1978, and the post-test in May, 1979, by six carefully trained substitute teachers. The Self Observation Scales and the Student Questionnaire were administered at both times during a single session in the students' classrooms. In addition, the My Opinion Survey was administered at the post-test during a second session. The students read the items to themselves (assisted by the administrator when necessary) except for the My Opinion Survey which was orally administered to grade 4 students. One make up session was held for students who were absent from the original session.

The confidentiality of the My Opinion Survey administration was enhanced by prelabeling questionnaires with student names on the cover sheet and students' school district identification numbers on page one. After receiving the questionnaire, students were instructed to tear off the cover page which displayed their names. In a prepared statement, administrators assured students of complete confidentiality and explained the need for identification numbers as a way of tracking students over time.

Teacher survey. The teacher pretest questionnaires were administered in September, 1978, during meetings at each school conducted by project staff. Teachers completed the post-test questionnaires in their classrooms while their students were being surveyed. To ensure the privacy of their responses, teachers were provided with questionnaires containing unique identifiers.

PROCESS EVALUATION METHODS AND FINDINGS

Descriptions of In-Service Training Sessions

To document the in-service training, a skilled observer attended all in-service sessions and made detailed notes about the content and process of the training. This observer later prepared lengthy descriptions of a) the agenda as planned and as actually followed, b) any materials used or distributed, c) the trainers' presentations, and d) the trainers' and teachers' verbal and nonverbal interactions.

The observer sat in the back of the room during training sessions and avoided interactions with the teachers. The trainers explained that the observer was present to "take notes so that we can develop a training manual."

The observations documented that the teachers enjoyed the sessions and participated enthusiastically in nearly all of the training activities. There was much open communication about both problems and successes in using the skills, and little criticism, conflict, or confrontation. Nearly all of the initial agendas for the sessions tended to underestimate the amount of time that the various activities would take, and thus, most of the sessions ran far behind schedule. Neither the trainers nor the participants seemed to mind this, however.

Teacher Attendance at Training Sessions

The first two of the twelve in-service sessions were mandatory for the 23 teachers (including one principal who taught half-time) potentially eligible to take the training, because these sessions were conducted during regular school time as part of two "minimum days" provided by the school district. The two mandatory sessions were attended by 21 and 20 teachers respectively.

Nineteen of the 23 potentially eligible teachers signed up for the ten voluntary in-service sessions. Of these teachers, 18 completed the training by attending at least nine of the ten sessions, and thereby earned \$200 stipends. Six of these teachers also attended a "reunion" session held two months after the end of training.

Other principals at the experimental schools were also invited to attend the training, and two of three attended frequently. The fourth principal attended only the first mandatory session.

Teacher Feedback on the Individual Training Sessions

At the end of each training session (except the final session), teachers were asked to complete a one-page "feedback form" anonymously. This form solicited ratings of, and comments about, the session. On the form, teachers rated each session for interest, organization, usefulness, and enjoyableness, using five-point rating scales with higher numbers signifying more positive ratings.

Teachers rated all sessions favorably. Of the 44 separate ratings (4 per session), 23 averaged above 4.50, and only 4 averaged less than 4.0 with none averaging less than 3.75. The teachers' high regard

for the training was also indicated by their frequent written compliments, and the rarity of any critical comments.

Interviews with Participating Teachers and Principals

Two of the experimental school principals and five of the participating teachers were interviewed individually by a member of the project's research team in April, 1979, approximately three months after the training had ended. Teachers were selected randomly for the interviews, which were semi-structured and open-ended. The interviews were tape-recorded with the respondents' permission, and lasted 10 to 25 minutes.

The interviews with teachers indicated that they thought very highly of the in-service training. All five felt that the training was enjoyable, useful, and relevant. All made a number of very positive general comments, such as "best I've had," "(we) need something like this every year," and "it was right down to gut level--what we needed." Critical comments about the training were rare, and three of the teachers could not offer any. One teacher thought the problem solving training "bogged down" at times. Another thought that the problem solving component was too complex, and that parts of the training were "too staged."

One of the principals evaluated the training as "very effective" and the other thought it had "increased (the teachers') level of awareness of some of the techniques." One thought the training could be improved by devoting more attention to techniques for dealing with

problem children; the other thought that the trainers' initial description of what was to be covered should have been more complete and clear.

The teachers differed in their reactions to the trainers' classroom visits. Two teachers found the visits useful and enjoyable because they got "good suggestions" and "lots of strokes." Two felt that the visits were unproductive because the trainers had not given them specific feedback and suggestions afterward. One teacher found the visits "not useful" because some of the trainer's feedback was based on a philosophy which she did not share, and because she did not want to change in the direction that the trainer suggested.

All five teachers reported that they were using all or most of the communication skills regularly. None of the teachers reported using the problem-solving skills regularly, because of the time required to implement these skills. Several also complained about the complexity of certain problem-solving skills, most notably force-field analysis.

Four teachers felt that their use of the skills had produced worthwhile changes in students. One said that her students were now more likely to make their own decisions. Another thought the training had helped him to establish better relationships with several students. A third reported that his students had begun to emulate his use of the communication skills. The fourth noted that her students were more serious about their school work, and wasted less class time.

The fifth teacher was not sure whether the training had produced any effect upon students.

Several of the teachers and both principals noticed changes in faculty behavior, including "smoother interactions between staff," better handling of crises or "semi-crises," "heightened awareness of even listening to each other," and much informal faculty discussion about the skills.

Questionnaire Surveys of Participating Teachers

At the completion of training (January, 1979), and again near the end of the school year (May, 1979), participating teachers provided detailed information about implementation by completing lengthy questionnaires. On the questionnaires, the teachers reported how often they used each in-service skill, rated the usefulness of each skill, and rated their own mastery of each skill.

Table 4 shows year-end data on frequency of skill implementation, and on usefulness and mastery of the skills. The first five skills listed in Table 4 are the communication skills; the second set of five are problem-solving skills; the last "skill"--self-concept enhancing activities--is a series of classroom exercises used apart from the teaching of the regular curriculum.

The data regarding frequency of implementation in Table 4 indicate that at year-end, nearly all of the teachers were using each of the five communication skills at least several times per week. More than two-thirds of the teachers reported using the self-concept enhancement activities on a weekly basis. However, the problem-solving skills were being used much less frequently; none of the teachers were using force-field analysis on a weekly basis, and approximately a third were

TABLE 4

TEACHERS' REPORTS ON FREQUENCY, UTILITY, AND QUALITY OF SKILL IMPLEMENTATION (N = 16)¹

24.

In-Service Skill	Percentage of Teachers Using the Skill at Least Several Times Per:		Mean Ratings of Value of the Skill (5=High Utility; 1=Low Utility)	Mean Ratings of Mastery of Skill ² (4=Excellent; 1=Poor)
	Month	Week		
Tailored Feedback	94 %	88 %	4.19	2.75
I-Messages	94	88	4.19	2.87
Clarifying Responses	100	75	4.25	2.94
Reflecting Feelings	100	94	4.31	3.00
Reflecting Content	100	88	4.06	2.94
Creating Problem Statements	75	38	3.69	2.69
Brainstorming	69	31	3.56	2.81
Evaluating Alternative Solutions	88	31	3.81	2.93
Force Field Analysis	47	0	3.07	2.54
Developing an Action Plan	53	33	3.56	2.64
Self-Concept Enhancement	94	69	4.38	3.06

¹Two teachers did not complete the questionnaire.²These ratings completed only by teachers who have used the skill in their classroom.

26

27

using the other problem solving skills weekly. This implementation pattern is consistent with the interview data reported above, and suggests that the teachers had integrated only the communication and self-concept enhancement skills into their general teaching styles.

With regard to the usefulness of the skills, Table 4 shows that, as might be expected from the frequency data, the teachers rated the problem solving skills as less useful in the classroom than the other skills. They tended to rate their own mastery of the skills as "good" rather than "excellent" or "fair," and to see themselves as least adept in using the problem solving skills.

To gather additional feedback on the in-service training, the mid-year questionnaire asked the teachers to evaluate the in-service course as a whole, using five-point rating scales. The teachers' ratings were very favorable with respect to interest ($\bar{M} = 4.58$), organization ($\bar{M} = 4.65$), usefulness ($\bar{M} = 4.37$), and enjoyableness ($\bar{M} = 4.53$). The teachers also rated the trainers highly along a number of dimensions. However, on the year-end questionnaire, the teachers gave the trainers' classroom follow-up visits an average usefulness rating of only 3.47 on a five-point scale, which is congruent with the mixed reviews that teachers gave the follow-up visits during the interviews.

To summarize, data from the questionnaire surveys indicate that teachers were using the communication and self-enhancement skills on a regular basis, were finding these skills useful, and considered themselves fairly adept at using these skills. However, they were not using the problem solving skills nearly as regularly, nor were they finding the

problem solving skills as useful. The teachers positively evaluated the overall training and the trainers' skills, but were less positive about the usefulness of the classroom follow-up visits.

Trainer Observations of Classroom Implementation

One of the trainers visited each classroom several times during the school year to observe the teachers applying the ECM skills. These observations ranged between 15 and 80 minutes, averaging 44.6 minutes ($SD = 14.68$). After each classroom observation, the trainer recorded a) the number of times each skill was used, and b) how well each skill was applied.

Averaged across teachers and across visits to the classrooms, the observational data show that teachers used some skills much more frequently than others. The skills that teachers used most often--better than twice per visit on average--were all communication skills: I-Messages ($M = 2.94$ implementations per visit; $SD = 1.70$), Tailored Feedback ($M = 2.87$; $SD = 1.29$), and Reflecting Content ($M = 2.56$; $SD = 1.91$). In contrast, none of the problem solving skills was used more than .22 times ($SD = .39$) per visit on average.

Although the teachers used some skills much more often than others, they were able to apply all of the skills with roughly equal competency. The trainer's average quality of implementation ratings were in the "B" to "B+" range for all skills, varying between 3.00 and 3.40 on a four-point scale, and averaging 3.26 ($SD = .24$) across all skills.

The teachers differed dramatically in how often they used the skills while being observed. At the extremes, one teacher used the skills an average of 19.4 times per observation, whereas another averaged only 3.80 times per visit. The average for all teachers combined was 11.22 ($SD = 3.85$) skill applications per visit.

The teachers differed much less in how well they implemented the skills. Here the range was between an average quality rating (across skills and over classroom visits) of 2.75 (i.e., "B-") for one teacher and 3.78 (i.e., "A-") for another. The average for all teachers was 3.26 ($SD = .24$). Interestingly, the quality ratings were not correlated significantly with how frequently the teachers used the skills while being observed (Spearman $r = .23$, $p < .20$).

These observational data are fairly consistent with the teacher self-report data described earlier, in that they show teachers using the communication skills regularly and the problem solving skills hardly at all. Furthermore, the observations indicate that teachers generally were using the skills in a competent and appropriate fashion.

Interviews with Control School Teachers and Nonparticipating Experimental School Teachers

To check for any diffusion of the treatment, interviews were conducted in April, 1979, with two principals and four teachers at control schools, and with one nonparticipating principal and two nonparticipating teachers at experimental schools. The semi-structured, open-ended interviews were tape-recorded with the respondents' permission.

None of the four control school teachers had heard anything about the content of the training or how it had been received in the experimental schools. Likewise, one of the two control school principals had not heard anything about the training or participants' reactions to it. Through conversations with friends at one of the experimental schools, the other control school principal knew which skills were being taught in the training, but had not made any effort to learn about the skills or to obtain similar training for teachers at his school.

One of the nonparticipating teachers had attended the two mandatory in-service sessions, and occasionally had heard other teachers talking about the training, but had made no effort to learn about the skills or to implement them in her classroom. The other nonparticipating teachers had not attended the mandatory sessions and had not heard anything about the training from others. The nonparticipating principal had attended the first mandatory session, but had not been "impressed that it was going to make that much change," and decided not to attend subsequent sessions.

In summary, neither the control nor nonparticipating teachers and principals knew much about the goals or techniques of the training, so that diffusion of the treatment is not considered a plausible threat to the validity of the outcome evaluation design.

OUTCOME RESULTS

A bi-level analysis approach was adopted for several reasons. The experimental design employed in this study involved random assignment of schools (and not classes or students) to the experimental or control condition. Least-squares analysis techniques assume the statistical independence of observations. Treating individual students' responses as independent when they share a common school and classroom environment (and common exposure to treatment) is problematic because the likelihood of obtaining a significant treatment effect spuriously is enhanced. In addition, student data and classroom aggregate data sometimes possess different substantive meanings and may be subject to different processes, resulting in conflicting findings. Thus, to complement the student-level analyses, we also conducted class-level analyses of the student data. We chose to examine class-level rather than school-level data because student interdependence occurs primarily within classes and not within schools. Teachers were responsible for implementation of the treatment, and they are assigned to particular classrooms.

Treatment of Missing Data

The problem of missing observations arose at two points in this analysis. In each instance, we adopted a strategy that allowed maximum use of the available data points while minimizing outcome bias.

Computation of Scale Scores. A scale score was computed for a student if at least 60% of the items comprising that scale were present. Any missing item score was replaced by the mean for that item in the appropriate subclass of the experimental design. This procedure utilized most of the item data and provided unbiased cell means. However, it constrained cell variances and inflated degrees of freedom artificially. When more than 40% of the items comprising a scale were missing, the student received a missing value for that scale.

Missing Scale Scores in Multivariate Analyses. Each of the multivariate analyses in this report was computed from summary data (i.e., subclass means and frequencies and a pooled within-class variance-covariance matrix). Choice of this input option enabled better utilization of data than typically is possible with available statistical routines, which exclude a case with even a single missing datum from analysis (c.f., MULTIVARIANCE: Finn, 1976; SPSS MANOVA: Cohen and Burns, Note 7). Specifically, the subclass mean and variance for a variable were based on all cases for which the variable was present, and the computation of each subclass covariance included every case in which the two variables were present. Hence, the number of observations contributing to covariances differed slightly within each subclass matrix. To compensate for our earlier inflation of degrees of freedom, we chose the minimum pairwise frequency to represent each subclass both in computing the pooled matrix and later for significance testing.

Analysis of Initial Equivalence and Attrition

An analysis was conducted to determine a) whether students in the experimental and control conditions were equivalent on the pre-test measures and b) whether the experimental and control conditions were affected differentially by attrition. Initial nonequivalence between treatment conditions has implications for both internal validity (the ability to make causal inferences about the effect of the treatment) and external validity (the ability to generalize findings to other populations). Attrition may affect internal validity if students missing from one condition differ systematically in kind from those missing from another condition. Attrition may also affect external validity if the missing (attrited) students differ systematically from those who remain in the sample (non-attrited).

As suggested by Jurs and Glass (1971), a two-way multivariate analysis of variance (MANOVA) was performed on the student pretest data. The factors of the design were treatment condition (experimental versus control), and attrition status (attrited versus non-attrited.) A main effect for treatment condition would provide evidence for initial nonequivalence (which is a threat to internal and external validity). A main effect for attrition status would provide evidence of limited external validity. The interaction between treatment condition and attrition status would indicate a threat to internal validity.

The main effect for treatment condition was significant, $F(6,871) = 3.18$, $p < .005$ providing evidence for initial nonequivalence of students. The scales affected were attitudes toward school, $F(1,876) = 10.19$,

$p < .002$, and attitudes toward peers, $F(1,876) = 6.93$, $p < .01$. For these two pretest scales, the experimental students had less positive attitudes than the control students. The mean differences between the conditions were .25 and .18, respectively, in terms of the metric of the pooled within-group standard deviations of the measures. The main effect for attrition status, $F(6,871) = 1.07$, and the interaction term, $F(6,871) = 1.47$, were not significant.

The initial nonequivalence between experimental and control students was analyzed further to determine whether it was consistent across grade and sex. This analysis and all subsequent student analyses were conducted separately for males and females because the two groups did not share a common variance-covariance structure. A two-way MANOVA was performed with treatment condition and grade level as the factors.

For the males, neither the multivariate main effect for treatment condition, $F(6,414) = 1.58$, nor the treatment by grade interaction, $F(12,828) = 0.72$, was significant. Thus, no evidence was found for initial differences. For the females, the multivariate treatment main effect, $F(6,397) = 2.20$, $p < .05$, was significant and the treatment by grade interaction, $F(12,794) = 1.12$, was not. The only scale affected significantly was attitudes toward school. For this measure, the treatment main effect, $F(1,402) = 6.95$, $p < .01$, and the treatment x grade interaction, $F(2,402) = 3.96$, $p < .02$, were significant. Examination of the treatment effects by grade level revealed a significant difference in grade 4, $t(140) = -3.71$, $p < .05$, but not in grade 5,

$t(126) = -1.04$, nor in grade 6, $t(136) = 0.22$. Experimental females in grade 4 scored .62 standard deviations lower than controls on the measure of attitudes toward school.

In summary, experimental and control male students were equivalent initially on the measured variables.. Experimental and control females initially differed, but this difference occurred primarily for a single measure, attitudes toward school, and in a single grade, grade 4. Experimental females in grade 4 were found to be less positive toward school prior to the treatment.

The pretest teacher data were examined for initial equivalence using univariate analyses of variance. Experimental and control teachers did not differ on initial measures of ECM objectives, $F(1,38) = 0.06$, or teacher satisfaction, $F(1,37) = 1.45$. However, experimental teachers scored .71 standard deviations lower than control teachers on faculty cohesiveness, $F(1,37) = 4.77$, $p < .04$. Experimental teachers were also found to have had more hours (29 vs. 8) of in-service training in the three years prior to the study, $F(1,37) = 4.77$, $p < .04$. Most notable was the fact that 64% of experimental teachers had prior in-service training in individualized instructional approaches as compared to only 17% of control teachers. The two groups of teachers did not reliably differ on 10 other background measures.

Multivariate Analyses of Student Data

For the student-level outcome analyses we chose to employ a two-level (experimental versus control) treatment condition variable. This approach ignores the distinction between the 75% of experimental students whose teachers participated in the ECM training and the 25% whose teachers did not participate. This analysis strategy provides a somewhat conservative estimate of the effect of the treatment on student outcomes as it assumes that all experimental group students were treated. The major advantages of this approach are a) that it preserves the experimental design, allowing stronger inferences to be made from the data, and b) that the obtained results may be generalizable to a wider range of situations.

For the multivariate analysis of the student data, separate analyses were performed on the male and female samples. A complete factorial design was employed for hypothesis testing, with treatment condition and grade level as the factors. Multivariate analysis of covariance (MANCOVA) was performed using pretest measures as covariates when the appropriate assumptions were met; otherwise, MANOVA was performed on post-test outcomes. Since the design was nonorthogonal due to unequal cell sizes, the reported main effect for each factor was tested after the variance associated with the other factor had been removed. The treatment x sex interaction was tested after removing the main effect variance. Effects were estimated subsequently using a reduced-rank model based upon the prior hypothesis testing. When significant interactions were obtained, the simple treatment effects were tested by

employing the Bonferroni t statistic (Myers, 1979). This statistic is a t -test that controls the Type I error rate for the number of tests performed. Because of space limitations, only results pertaining to the treatment condition and treatment interaction are discussed.

The post-test measures were divided into three subsets based upon *a priori* hypotheses about their interrelationships and empirical findings concerning the homogeneity of their covariance structures.⁷ One subset (A) consisted of the affective measures: locus of control: success (Control Suc), locus of control: failure (Control Fail), social self-esteem (Social Self), attitudes toward peers (Att Peers), affective teaching climate (Affec Climate), attitudes toward school (Att Sch), academic self-esteem (Acad Self), and perceived peer attitudes toward school (Peer Att Sch). A second subset (B) included the two attendance measures, unexcused absences (Unexc Abs) and total absences (Total Abs), and the achievement measures, reading achievement (Read) and mathematics achievement (Math). The final subset (C) consisted of the drug measures: perceived costs of alcohol (Alc Costs), cigarettes (Cig Costs), and marijuana use (Pot Costs), and involvement in alcohol (Alc Involve), cigarettes (Cig Involve) and marijuana (Pot Involve) use. The teacher ratings of student minor (Minor) and major (Major) misbehavior were not analyzed at the student level due to heterogeneity of variance and covariance.

⁷Due to heterogeneity of variance, the following measures were transformed [$\log (x + 1)$]: Unexc Abs, Total Abs, Cig Involve, Pot Involve, and Major.

Each of the three subsets of post-test measures was analyzed separately for each sex. Preliminary analyses⁸ were performed to identify the set of pretest outcomes that best predicted each set of post-test outcomes and did not violate the assumptions of the covariance analysis (primarily, that the regression hyperplanes be homogeneous across the cells of the design). For Set A, two covariate composites were computed, a student composite (Affec Comp) and a teacher composite (Teach Comp). Affec Comp consisted of a unit weighted combination of student pretest measures: Control Suc, Control Fail, Social Self, Att Peers, Affec Climate, and Att Sch. Teach Comp included two teacher pretest measures, the amount of in-service training in the prior three years minus the amount of teaching experience.⁹ For sets B and C, covariate selection was abandoned because either the pretest measures were not good predictors of the post-test outcomes or the homogeneity assumptions (for covariance and regression slopes) were violated. For these sets of measures, MANOVA results have been reported.

⁸Principal component analyses were conducted on the pretest data and cononical correlation analyses on the pretest and post-test data.

⁹For females the regression slopes were not parallel; thus, the analysis was also computed with the treatment x covariate terms in the model (Reichardt, 1979). The results from this model did not differ from the reported results.

The student data have been summarized in Table 5 and the results of the multivariate analyses have been summarized in Table 6. For the males, no significant treatment-related effects were obtained. For the females, A significant multivariate treatment main effect was obtained on set C, $F(6,449) = 2.24$, $p < .04$. No treatment related effects were obtained on sets A or B.

For females, reliable univariate treatment main effects occurred for two of the measures in set C: Pot Involve, $F(1,454) = 8.38$, $p < .004$, and on Pot Costs, $F(1,454) = 5.34$, $p < .03$. Experimental females scored .27 and .22 standard deviations lower¹⁰ than controls on these measures of involvement in, and perceived costs of, marijuana use. Secondary analyses were performed on the individual items from which the marijuana involvement measure had been constructed: perceived peer attitudes and use, attitudes, life use and current use. A significant treatment main effect was obtained on the perceived peer attitudes measure, $F(1,459) = 7.24$, $p < .008$, and the perceived peer use measure, $F(1,459) = 6.18$, $p < .02$. In addition, a significant treatment x grade interaction was obtained on the latter measure, $F(2,459) = 4.15$, $p < .02$. Experimental females were .25 standard deviations lower than controls on the measure of perceived peer attitudes toward marijuana. Experimental females in grade 5 were .62 standard deviations lower than controls, $t(147) = -3.74$, $p < .05$, on the measure of perceived peer use of marijuana. Reliable treatment differences were not obtained in grades 4 and 6 for this measure.

¹⁰All drug related scales were scored so that a higher score reflected a more "pro-drug" position.

TABLE 5

MEANS AND STANDARD DEVIATIONS FOR STUDENT DATA
BY SEX BY TREATMENT CONDITION FOR GRADE 4

Measure	Test	Male Students				Female Students			
		Experi- mental		Control		Experi- mental		Control	
		M	SD	M	SD	M	SD	M	SD
Affec Comp	pre	1.70	.16	1.76	.14	1.73	.14	1.78	.13
Control Suc	pre	1.76	.18	1.78	.19	1.83	.15	1.83	.15
Control Suc	post	1.80	.20	1.77	.25	1.88	.15	1.86	.18
Control Fail	pre	1.59	.22	1.64	.23	1.62	.21	1.65	.25
Control Fail	post	1.68	.26	1.66	.25	1.71	.24	1.71	.24
Social Self	pre	1.69	.27	1.76	.21	1.69	.24	1.72	.25
Social Self	post	1.77	.28	1.78	.21	1.76	.25	1.76	.26
Att Peers	pre	1.67	.28	1.75	.24	1.66	.31	1.73	.24
Att Peers	post	1.69	.31	1.76	.25	1.70	.30	1.72	.29
Affec Climate	pre	1.85	.24	1.90	.16	1.89	.16	1.93	.17
Affec Climate	post	1.71	.20	1.73	.19	1.74	.17	1.78	.15
Att Sch	pre	1.63	.30	1.75	.24	1.64	.31	1.80	.22
Att Sch	post	1.64	.31	1.70	.30	1.69	.26	1.78	.26
Acad Self	post	1.73	.25	1.70	.25	1.76	.23	1.76	.22
Peer Att Sch	post	1.70	.28	1.72	.25	1.74	.23	1.79	.20
Unexc Abs	post	.24	.29	.23	.31	.23	.30	.23	.31
Total Abs	post	.18	.16	.20	.17	.25	.19	.23	.20
Read	post	5.41	1.67	5.84	1.55	6.13	1.39	5.91	1.56
Math	post	5.65	1.72	5.61	1.64	6.10	1.40	5.88	1.60
Minor	post	1.95	1.06	2.18	1.23	1.51	.69	1.37	.71
Major	post	.39	.15	.31	.04	.34	.11	.31	.03
Alc Costs	post	.47	.53	.56	.44	.51	.41	.44	.42
Cig Costs	post	.30	.35	.42	.39	.45	.38	.41	.38
Pot Costs	post	.31	.37	.41	.43	.39	.39	.40	.36
Alc Involve	post	1.35	.69	1.49	.67	1.26	.53	1.15	.55
Cig Involve	post	.24	.05	.27	.07	.26	.07	.24	.06
Pot Involve	post	.23	.04	.24	.09	.23	.05	.23	.04

TABLE 5 (CONT.)

39.

MEANS AND STANDARD DEVIATIONS FOR STUDENT DATA
BY SEX BY TREATMENT CONDITION FOR GRADE 6

Measure	Test	Male Students				Female Students			
		Experi- mental		Control		Experi- mental		Control	
		<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Affec Comp	pre	1.78	.15	1.79	.13	1.78	.15	1.81	.12
Control Suc	pre	1.82	.17	1.80	.18	1.84	.14	1.83	.15
Control Suc	post	1.87	.19	1.78	.23	1.90	.12	1.88	.15
Control Fail	pre	1.73	.22	1.71	.22	1.73	.19	1.78	.18
Control Fail	post	1.81	.21	1.69	.26	1.80	.20	1.82	.21
Social Self	pre	1.83	.22	1.86	.20	1.79	.25	1.83	.20
Social Self	post	1.86	.18	1.84	.22	1.84	.23	1.86	.20
Att Peers	pre	1.73	.26	1.80	.22	1.67	.33	1.76	.23
Att Peers	post	1.77	.26	1.73	.28	1.74	.27	1.76	.26
Affec Climate	pre	1.90	.18	1.90	.17	1.91	.14	1.92	.15
Affec Climate	post	1.78	.15	1.69	.22	1.79	.13	1.77	.16
Att Sch	pre	1.66	.29	1.67	.27	1.75	.23	1.74	.22
Att Sch	post	1.68	.28	1.57	.28	1.71	.26	1.69	.27
Acad Self	post	1.75	.24	1.71	.25	1.77	.26	1.77	.21
Peer Att Sch	post	1.83	.21	1.70	.25	1.81	.21	1.73	.24
Unexc Abs	post	.17	.29	.25	.29	.17	.28	.23	.29
Total Abs	post	.21	.18	.24	.17	.25	.18	.23	.20
Read	post	6.13	1.54	5.88	1.77	6.33	1.67	6.45	1.68
Math	post	6.04	1.52	5.56	1.82	6.02	1.52	5.98	1.67
Minor	post	2.16	1.39	1.81	1.23	1.50	.88	1.12	.37
Major	post	.35	.10	.34	.12	.31	.04	.30	.02
Alc Costs	post	.49	.53	.63	.58	.43	.48	.47	.42
Cig Costs	post	.48	.51	.51	.44	.47	.45	.57	.46
Pot Costs	post	.43	.53	.53	.50	.38	.49	.48	.43
Alc Involve	post	1.50	.73	1.77	.71	1.38	.65	1.45	.55
Cig Involve	post	.29	.10	.31	.10	.30	.10	.33	.09
Pot Involve	post	.26	.10	.28	.11	.25	.09	.27	.07

TABLE 5 (CONT.)

40.

MEANS AND STANDARD DEVIATIONS FOR STUDENT DATA
BY SEX BY TREATMENT CONDITION FOR GRADE 5

Measure	Test	Male Students				Female Students			
		Experi- mental		Control		Experi- mental		Control	
		<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Affec Comp	pre	1.73	.14	1.74	.13	1.78	.14	1.78	.13
Control Suc	pre	1.78	.19	1.80	.15	1.86	.15	1.85	.13
Control Suc	post	1.84	.18	1.80	.21	1.86	.17	1.85	.20
Control Fail	pre	1.61	.20	1.59	.21	1.68	.22	1.71	.23
Control Fail	post	1.69	.24	1.70	.23	1.80	.23	1.78	.24
Social Self	pre	1.79	.23	1.80	.22	1.75	.26	1.74	.24
Social Self	post	1.83	.22	1.84	.19	1.84	.24	1.83	.23
Att Peers	pre	1.70	.28	1.73	.25	1.75	.27	1.73	.25
Att Peers	post	1.73	.29	1.73	.28	1.74	.28	1.72	.28
Affec Climate	pre	1.87	.20	1.86	.23	1.93	.15	1.92	.17
Affec Climate	post	1.74	.18	1.69	.19	1.81	.11	1.75	.16
Att Sch	pre	1.63	.29	1.66	.26	1.73	.26	1.77	.23
Att Sch	post	1.66	.29	1.61	.29	1.74	.26	1.72	.30
Acad Self	post	1.75	.24	1.75	.22	1.80	.22	1.69	.28
Peer Att Sch	post	1.75	.23	1.66	.26	1.85	.18	1.72	.27
Unexc Abs	post	.23	.30	.27	.33	.25	.30	.25	.33
Total Abs	post	.23	.20	.23	.18	.24	.18	.22	.19
Read	post	6.05	1.49	5.95	1.58	6.14	1.51	5.77	1.48
Math	post	5.99	1.47	5.70	1.77	5.87	1.63	5.72	1.56
Minor	post	2.26	1.78	2.63	1.51	1.41	.79	1.98	1.31
Major	post	.34	.09	.34	.11	.31	.06	.33	.10
Alc Costs	post	.55	.53	.59	.44	.41	.40	.43	.42
Cig Costs	post	.42	.46	.39	.36	.40	.39	.50	.47
Pot Costs	post	.38	.50	.43	.46	.25	.29	.40	.37
Alc Involve	post	1.47	.73	1.53	.61	1.11	.42	1.28	.59
Cig Involve	post	.27	.09	.27	.09	.25	.06	.29	.09
Pot Involve	post	.24	.08	.25	.08	.21	.03	.25	.07

TABLE 6

SUMMARY OF MULTIVARIATE ANALYSES OF STUDENT DATA

<u>Source of Variation</u>	<u>Subset of Measures</u>	<u>Males</u>			<u>Females</u>		
		<u>df</u>	<u>Multi-variate F</u>	<u>p <</u>	<u>df</u>	<u>Multi-variate F</u>	<u>p <</u>
Treatment Condition (T)	A ^a	8,375	1.38	ns	8,366	1.26	ns
	B ^b	4,428	2.07	ns	4,417	0.87	ns
	C ^c	6,458	1.50	ns	6,449	2.24	.04
Grade (G)	A	16,750	0.82	ns	16,732	2.55	.001
	B	8,856	1.73	ns	8,834	1.61	ns
	C	12,916	2.76	.002	12,898	6.09	.001
T x G	A	16,750	1.07	ns	16,732	1.49	ns
	B	8,856	0.78	ns	8,834	0.44	ns
	C	12,916	0.62	ns	12,898	1.20	ns

^aSubset A analysis: Variates were Control Suc, Control Fail, Social Self, Att Peers, Affec Climate, Att School, Acad Self, Peer Att Sch. The covariate was Affec Comp.

^bSubset B analysis: Variates were Unexc Abs, Total Abs, Read, and Math.

^cSubset C analysis: Variates were Alc Costs, Cig Costs, Pot Costs, Alc Involve, Cig Involve, Pot Involve.

Analyses of Class-Level Student Data

The class-level student analysis was performed on unique class, sex and grade-level units. For example, a combined 4-5 grade class contributed four units to the analysis, each composed of the appropriate set of students from the class: grade 4 males, grade 4 females, grade 5 males, and grade 5 females. For each analysis unit the mean scores were computed for all student measures. Separate analyses were then conducted on males and females. With this approach there is greater interdependence among units of analysis than in a strictly class-level or school-level analysis; however, the approach enables control for, and examination of, the association of student grade and sex with the treatment effects.

Univariate analyses of covariance were applied to each post-test measure with treatment condition and grade level as factors in the design and Affec Comp as the covariate.¹¹ When the assumptions of the analysis were violated, analysis of variance was performed on the post-test data.

Results of the univariate analyses of variance and covariance on the class-level male and female student data are summarized in Table 7. For males, the class-level results were consistent with the student-level results for 11 of the 18 variables: Social Self, Att Peers, Acad Self,

¹¹ Teach Comp did not predict post-test outcomes systematically at the class level, and hence was not used as a covariate.

TABLE 7

SUMMARY OF UNIVARIATE ANALYSES OF CLASS-LEVEL STUDENT DATA

Measure	MALES							FEMALES						
	Analysis Type ^a	Treatment			Treatment x Grade			Analysis Type	Treatment			Treatment x Grade		
		df	F	p <	df	F	p <		df	F	p <	df	F	p <
Control Suc	C	1,38	17.41	.001	2,38	0.62	ns	A	1,39	1.15	ns	2,39	0.09	ns
Control Fail	C	1,38	5.52	.03	2,38	2.89	ns	A	1,39	0.07	ns	2,39	0.24	ns
Social Self	C	1,38	1.37	ns	2,38	0.12	ns	C	1,38	0.22	ns	2,38	0.37	ns
Att Peers	C	1,38	0.11	ns	2,38	0.62	ns	C	1,38	0.83	ns	2,38	0.15	ns
Affec Climate	A	1,39	4.90	.04	2,39	1.93	ns	A	1,39	2.08	ns	2,39	1.89	ns
Att Sch	C	1,38	7.11	.02	2,38	1.25	ns	A	1,39	0.19	ns	2,39	0.49	ns
Acad Self	C	1,38	2.63	ns	2,38	1.46	ns	C	1,38	8.66	.01	2,38	0.64	ns
Peer Att Sch	C	1,38	10.23	.003	2,38	0.20	ns	A	1,39	3.34	ns	2,39	2.24	ns
Unexc Abs	A	1,39	1.32	ns	2,39	0.89	ns	A	1,39	0.24	ns	2,39	0.42	ns
Total Abs	A	1,39	0.84	ns	2,39	0.38	ns	A	1,39	0.00	ns	2,39	0.09	ns
Read	C	1,38	0.60	ns	2,38	0.21	ns	C	1,38	0.51	ns	2,38	0.32	ns
Math	C	1,38	2.73	ns	2,38	0.21	ns	C	1,38	2.03	ns	2,38	0.40	ns
Minor	A	1,34	0.02	ns	2,34	0.33	ns	A	1,35	0.00	ns	2,35	2.66	ns
Major	C	1,33	0.05	ns	2,33	0.42	ns	A	1,35	0.03	ns	2,35	1.12	ns
Alc Costs	A	1,39	4.38	.05	2,39	0.19	ns	A	1,39	0.28	ns	2,39	0.19	ns
Cig Costs	A	1,39	0.59	ns	2,39	0.83	ns	A	1,39	2.59	ns	2,39	1.03	ns
Pot Costs	A	1,39	2.38	ns	2,39	0.02	ns	A	1,39	3.98	ns	2,39	0.79	ns
Alc Involve	A	1,39	4.36	.05	2,39	0.17	ns	A	1,39	1.05	ns	2,39	0.41	ns
Cig Involve	A	1,39	1.89	ns	2,39	0.06	ns	A	1,39	3.18	ns	2,39	1.19	ns
Pot Involve	A	1,39	1.63	ns	2,39	0.13	ns	A	1,39	5.47	.03	2,39	1.61	ns

^aC = analysis of covariance

V = analysis of variance

Unexc Abs, Total Abs, Read, Math, Cig Costs, Pot Costs, Cig Involve, and Pot Involve. On these measures both levels of analysis yielded no evidence for treatment effects. For seven measures--Control Suc, Control Fail, Affec Climate, Att Sch, Peer Att Sch, Alc Costs, and Alc Involve--the class-level analyses revealed positive treatment effects, whereas the student analyses did not. For five of these measures, Control Suc, Control Fail, Affec Climate, Att Sch, and Peer Att Sch, the discrepant results are most likely due to differences in the models specified in the two analyses. The student-level analysis on the students' post-test scores was adjusted for the differential effect of their teachers' prior in-service training (in addition to student affective pretest differences). The class-level analysis did not control for the initial bias attributable to prior in-service training and is therefore a less adequate model. The positive treatment effects found for Alc Costs and Alc Involve in the class-level analysis were partially replicated in the student analysis if the nonsignificant multivariate treatment main effect is ignored, and the respective univariate statistics for these two measures is examined, $F(1,463) = 3.68$, $p < .06$, and $F(1,463) = 6.66$, $p < .01$. Experimental students were .18 and .24 standard deviations lower than control students on these measures. However, considering the initial teacher differences and lacking appropriate pretest measures of these variables, attributing these differences to the treatment is risky. In addition, it is not in keeping with standard practice to ignore the multivariate results.

No effects were found on Minor and Major, the two measures that were analyzed only at the class level.

For females, results from the class-level analyses were consistent with those from the student-level analyses for 16 of the 18 variables (all measures except Acad Self and Pot Costs). A consistent positive treatment effect was found for one measure, Pot Involve, and no treatment-related effects were obtained for the 15 others. A positive treatment effect was obtained for Acad Self in the class-level analysis but not in the student-level analysis. The different results are probably due to the different models specified. The student-level analysis adjusted for preexisting treatment condition differences on Teach Comp while the class-level analysis did not. For Pot Costs the class-level analysis revealed no treatment effect, whereas the student-level analysis found an effect. However, the class-level analysis did approach significance, $F(1,39) = 3.34$, $p < .10$, which provides partial support for a positive treatment effect. No effects were found on Minor and Major, the two measures analyzed only at the class level.

In sum, the student- and class-level analyses generally produced consistent results. Where results were inconsistent, the differences can be explained in terms of the the different models specified or the statistical power of the tests. For males, no consistent evidence was obtained that indicated a positive treatment effect. For females, a positive treatment effect was found consistently on

the measure of involvement in marijuana use. However, this obtained effect may be due to an improperly specified model, attributable to the lack of pretest measures that better specify the preexisting differences.

Analysis of Teacher Data

One-way (treatment condition) univariate analyses of covariance were applied to each teacher post-test measure with the corresponding pretest used as a covariate. The results are summarized in Table 8. No treatment effect was found on any of the measures.

The Relationship of Student Outcomes to Treatment Implementation

The analyses discussed thus far compared treated versus untreated students, teachers, and classrooms. Just as it is reasonable to expect these groups to differ on the basis of presence or absence of the ECM treatment, it is possible that treated classes might differ among themselves as a function of the "amount" of treatment they received. Accordingly, analyses were conducted to investigate the relationships between the experimental classrooms' average student post-test outcomes and their teachers' levels of treatment implementation. A treatment implementation measure was constructed from three sources: a) the trainer's observations of how frequently the teachers' implemented the skills; b) the teachers' self-reports on how frequently they implemented each skill; and c) the teachers' responses to the classroom behavior inventory. The pattern of intercorrelations among these three sources of data provided evidence for the convergent and discriminant validity (Campbell and Fiske, 1959) for measures of three

TABLE 3

SUMMARY OF UNIVARIATE ANALYSES OF TEACHER DATA

<u>Measure</u>	<u>df</u>	<u>Treatment</u>	
		<u>F</u>	<u>p <</u>
Teacher Satisfaction	1,29	0.47	ns
Faculty Cohesiveness	1,28	1.14	ns
ECM Objectives	1,29	1.22	ns

skills: 1) I-messages, a communications skills, based upon items from all three sources; 2) reflecting content, a listening skill, based upon items from b and c; and 3) self-concept enhancement activities, based upon items from b and c. Each teacher's scores on the above measures were standardized and summed to produce the treatment implementation measure.

The measure of treatment implementation was correlated with each of the class mean student post-test outcome measures. In addition, the implementation measure was dichotomized at the median resulting in two groups of experimental teachers, a high implementation group and a low implementation group. Analyses of variance were performed on the class mean student pretest data to determine whether high and low implementation classes initially differed on the pretest measures. The results indicated no initial differences between the two groups for either the males or the females. Since no initial differences were obtained, univariate analyses of variance were performed on the class mean student post-test data with implementation level (high versus low implementation) as the independent variable.¹² The results from the analyses for the males and females have been summarized in Table 9. No significant differences were obtained, indicating that the level of treatment implementation did not have a detectable effect upon any student post-test outcome for either males or females. Of course, the statistical power of these analyses was limited due to the small number of classes that were available ($N = 14$).¹³

¹²Both the median split ANOVA and the correlational analysis were performed because each analysis makes different assumptions about the relationship between treatment implementation and student outcomes.

¹³Two classes were eliminated from the analysis due to missing treatment implementation scores.

TABLE 9

SUMMARY OF TREATMENT IMPLEMENTATION ANALYSES ON CLASS-LEVEL STUDENT DATA

Measure	MALES		FEMALES	
	<u>r(12)</u>	<u>F(1,12)</u>	<u>r(12)</u>	<u>F(1,12)</u>
Control Suc	.03	0.02	.43	1.51
Control Fail	.52	1.01	.29	2.51
Social Self	-.25	2.56	-.15	0.88
Att Peers	.31	2.26	.02	0.01
Affec Climate	-.02	0.32	-.05	0.28
Att Sch	.01	1.81	-.24	1.95
Acad Self	-.01	0.29	-.07	2.61
Peer Att Sch	.11	0.73	-.34	0.20
Unexc Abs	-.22	0.65	.31	1.04
Total Abs	-.09	0.37	-.06	0.46
Read	.05	0.16	.04	1.28
Math	.07	0.05	.34	0.10
Minor	.00	0.37	-.44	0.93
Major	-.10	0.95	-.31	2.48
Alc Costs	.08	0.34	-.25	1.14
Cig Costs	.21	1.60	-.33	2.01
Pot Costs	.20	1.22	-.07	0.52
Alc Involve	.30	1.32	.00	0.13
Cig Involve	.17	0.92	.04	0.16
Pot Involve	.26	1.35	.24	0.30

DISCUSSION

Process evaluation data indicated that the teachers in the experimental group found the ECM in-service training to be interesting, enjoyable, well-organized, and useful. The teachers regularly attended the training, and enthusiastically participated in it. They were divided, however, about the usefulness of the trainer's classroom follow-up visits; most saw these as valuable but a few evaluated the visits as unproductive.

Although most teachers used most of the ECM skills and self-esteem enhancement activities with some regularity in their classrooms, few used the problem solving skills more than sporadically. Apparently, the teachers found the problem solving skills too complicated and time-consuming for easy application. There was considerable variation among the teachers in how often they used the ECM skills, although they were able to use all of the skills with at least moderate proficiency (including the problem solving skills).

None of the expected treatment effects on teachers was found; experimental and control group teachers did not differ significantly with respect to faculty cohesiveness, job satisfaction, or self-evaluated effectiveness at achieving the ECM objectives. The absence of significant effects on these variables is especially disappointing and perplexing in light of the teachers' reports that the skills improved their teaching effectiveness and the training brought them closer to their colleagues.

Nor is the lack of significant teacher differences explained by the conservative decision to include the five nonparticipating teachers from the experimental schools in the treatment group for purposes of making comparisons with the control teachers; results were not materially affected by discarding the data from these five teachers.

Regarding effects on students, ECM showed no consistently significant impact on experimental versus control males across both the individual- and class-level student outcome analyses. The class-level analyses did show positive treatment effects on seven outcome variables, but none of these was significant on the student-level analysis. The discrepancies between the two types of analysis are best explained by the fact that the student-level analysis controlled for sizeable between-group teacher differences as well as between-group student differences, whereas the class-level analysis controlled only for student-level differences. Thus, the student-level analysis better specified preexisting student and teacher differences, and compels the conclusion that ECM had no significant effects on male.

Both student- and class-level outcome analyses for experimental versus control females showed significantly lower involvement in marijuana use. There was also evidence that experimental females perceived greater adverse consequences from marijuana use. However, no pretest data on these variables were collected, so that any preexisting differences between the groups on the variables could not be specified in the analyses. Thus, the positive effects may be artifacts arising from inadequate specification of preexisting student and teacher differences.

An exploratory analysis of the effects of experimental teachers' implementation rates upon student outcomes showed no significant relationships. However, the analysis had very limited statistical power due to the small number of available classes.

Taken together, the above pattern of findings suggests that ECM had few significant effects upon teachers and students during the first year of its delivery. This may be in part due to an inadequate in-service training design and inadequate implementation of the ECM skills--none of the experimental teachers used the problem solving skills consistently, and there was much variation among the teachers in use of the other skills. (ECM has been revised for delivery in subsequent years; most notably, the problem solving skills have been replaced by classroom discipline skills.) It is also possible that the duration of the intervention--five or six months at best--was insufficient to produce the desired effects, and that continued implementation of the ECM skills through one or two additional school years will prove more consequential.

ECM's true effects may have been obscured by lack of certain pretest data, insensitive outcome measures, or a weak research design. Pre-testing did show substantial differences between experimental and control group teachers, and between female students in the experimental versus control groups. Although covariance analyses were used where appropriate and feasible, these initial non-equivalencies and the lack of pretest data on certain variables may have prevented detection of certain effects.

True effects almost certainly were not obscured by the conservative decision to include nonparticipating teachers from the experimental schools and their students in the treatment groups. Separate analyses of the data

for these teachers and their students showed that their inclusion did not affect materially the comparisons between experimental and control teachers and students.

Perhaps the most appropriate statement about ECM that can be made at this time is that another year of implementation and data collection are necessary before a preliminary judgment should be made about the efficacy of this strategy.

REFERENCE NOTES

1. Michigan Department of Education & The Office of Substance Abuse Services, Department of Public Health. 1976-77 Evaluation Report: Substance Abuse Prevention Education Program, undated.
2. Georgia Department of Human Resources & Georgia State Department of Education. Life Skills for Mental Health, 1977.
3. Bandt, P., Hammond, M., Wisdo, T., & Mitzel, H. A Follow-Up Study of Teachers Trained in Affective Classroom Approaches to Primary Prevention of Drug and Alcohol Abuse. State College, PA: Addictions Prevention Laboratory, 1976.
4. Stenner, A. & Katzenmeyer, W. Technical Manual and User's Guide: Self-Observation Scales, NTS Research Corporation, Durham, NC: undated.
5. Moskowitz, J., Condon, J., Brewer, M., Schaps, E. & Malvin, J. The Napa Project: Scaling of Student Self-Report Instruments. Prevention Branch, National Institute on Drug Abuse, 1979.
6. Kaufman, M., Semmel, M. & Agard, J. *Supplemental Materials to Year I Interim Report, Part II*. Washington, D. C.: U. S. Office of Education, Project Prime, undated.
7. Cohen, E. and Burns, P. SPSS--MANOVA--Multivariate Analyses of Variance and Covariance. Northwestern University, Vogelback Computing Center, Document No. 413 (A), 1977.

REFERENCES

- Baskin, E. & Hess, R. Does affective education work? A review of seven programs. *Journal of School Psychology*, 1980, 18, 40-50.
- Campbell, D. and Fiske, D. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, 56, 81-105.
- Crandall, V., Katkovsky, W. & Crandall, V. Children's belief in their own control of reinforcements in intellectual-academic achievement situations. *Child Development*, 1965, 36, 91-109.
- Finn, J. *MULTIVARIANCE: Univariate and Multivariate Analysis of Variance, Covariance and Regression*. Ann Arbor, Michigan: National Educational Resources, Inc., 1972.
- Halpin, A. & Croft, D. *The Organizational Climate of Schools*. Chicago: Midwest Administration Center, University of Chicago, 1963.
- Instructional Objectives Exchange. *Attitude Toward School K-12*. (rev. ed.) Los Angeles: Instructional Objectives Exchange, 1972 a.
- Instructional Objectives Exchange. *Knowledge and Attitudes of Drug Usage, Grades 4-12*. Los Angeles: Instructional Objectives Exchange, 1973.
- Instructional Objectives Exchange. *Measures of Self-Concept K-12*. (rev. ed.) Los Angeles: Instructional Objectives Exchange, 1972 b.
- Jurs, S. & Glass, G. The effect of experimental mortality on the internal and external validity of the randomized comparative experiment. *Journal of Experimental Education*. 1971, 40, 62-66.
- Madden, R., Gardner, E., Rudman, H., Karlson, B. & Merwin, J. *Stanford Achievement Tests: Norms Booklet (Primary Level III)*. New York: Harcourt Brace Javanovich, 1973.
- Myers, J. *Fundamentals of Experimental Design (Third Edition)*. Boston: Allyn and Bacon, Inc., 1979.
- Reichardt, C. The statistical analysis of data from nonequivalent control group designs. In T. Cook and D. Campbell, *Quasi-experimentation. Design and Analysis Issues for Field Settings*, Chicago: Rand McNally, 1979.
- Rempel, A. & Bentley, R. The measurement of teacher morale: a factor analysis approach. *Educational and Psychological Measurement*, 1964, 24, 631-643.
- Schaps, E., DiBartolo, R., Moskowitz, J., Palley, C., & Churgin, S. A review of 127 prevention program evaluations, *Journal of Drug Issues*, in press.